

## CÔNG CỤ THỐNG KÊ MIỄN PHÍ TRÊN MẠNG

Rất nhiều thông tin thống kê sẵn có miễn phí trên mạng. Thông tin gồm dữ liệu và các tập dữ liệu và cũng có cả các sách giáo khoa về thống kê, các danh sách thư điện tử, phần mềm và nhiều địa chỉ về các chủ đề đặc biệt, như dịch tễ học, dự báo, trình bày dữ liệu, biên tập dữ liệu, cây số liệu (multiple imputation) và phân tích điểm

khuyến hướng (propensity score analysis). Sau đây là một số địa chỉ hữu ích:

Địa chỉ tốt nhất để bắt đầu học thống kê học là HyperStatistics Online, ở địa chỉ <http://davidmlane.com/hyperstat/index.html>. Đây là địa chỉ tốt nhất vì ngoài việc là một cuốn sách về thống kê học, đây còn có một danh sách toàn diện các cuốn sách

về thống kê học khác trực tuyến. Đa số sách này là rất cơ bản cho bậc đại học. Có một cuốn văn bản Statsoft, <http://www.statsoft.com/textbook/stathome.html> có các chủ đề bậc cao.

Sau khi phân tích các dữ liệu, vấn đề rất cần thiết là phải biết cách trình bày tốt nhất các kết quả. Các địa chỉ tốt nhất là: Conventions for the Graphical Display of data <http://www.nmsu.edu/techprof/backgrnd/graphunit/graphback.htm>, và Washington Statistical Practice Society Methodology Seminars. Data Presentation: A Guide To Good Graphics <http://www.science.gmu.edu/~wss/methods/zawitzg.html>. Còn Guide to Good Statistical Practice của Cục Thống kê vận tải ở địa chỉ <http://www.bts.gov/programs/statpol/btsguide.html> có một phần hữu ích về trình bày các kết quả.

Còn có rất nhiều phần mềm miễn phí trên mạng. Nơi tốt nhất để tìm phần mềm thống kê miễn phí là Free Statistical Software địa chỉ <http://members.aol.com/johnp71/javasta2.html>. Địa chỉ này liệt kê danh sách phần mềm phục vụ mục đích chung cũng như phần mềm dành cho các mục đích đặc thù, như đường hồi quy, dịch tễ học, điều tra và lập trình. Cũng có mô tả ngắn gọn từng mảng.

Một địa chỉ chung, đặc biệt giúp cho các sinh viên, là American Statistical Association <http://www.amstat.org/>. Có một danh sách kết nối rất tốt, bao gồm các hội thống kê, các nguồn điện tử và các cơ quan tài trợ. Cũng có một địa chỉ việc làm. Những ai quan tâm đến các cuộc điều tra, hãy tìm phần về các phương pháp nghiên cứu điều tra ở

<http://www.amstat.org/sections/srms/>, đặc biệt là các xê - ri What is a Survey.

Địa chỉ tốt nhất để bắt đầu các danh sách thư điện tử là Allstat, ở <http://www.ltsn.gla.ac.uk/allstat>. Đây là địa chỉ tốt nhất vì có một danh sách thư điện tử rất thú vị, đó là một danh sách đầy đủ các danh sách thư điện tử thống kê khác. Có lẽ danh sách thống kê chung phổ biến nhất là stat-l, ở <http://www.cmh.edu/stats/faq/faq.htm>. Một danh sách hữu ích khác, không ở Allstat, là Epidemio, ở <http://mapageweb.umontreal.ca/philipp/> (đặt mua ở LISTPROC@CC.UMONTREAL.CA, gửi "SUBSCRIBE EPIDEMIO-L, họ tên"). Danh sách này nói về dịch tễ học.

Nói chung, nơi tốt nhất để xem nguồn dữ liệu là Statistical Resources on the Web <http://www.lib.umich.edu/govdocs/stats.html>. Đây là một địa chỉ dẫn toàn diện các dữ liệu về nhiều chủ đề, gồm sức khoẻ, nhân khẩu học, lao động, kinh tế học, môi trường và nhiều thứ khác nữa. Một địa chỉ khác là Statistics, ở <http://www.statistics.com/>. Một nơi xuất phát cho các dữ liệu xã hội, chính trị và kinh tế là trang dữ liệu Social Change <http://gsciology.icaap.org/data.html>, cũng nối với một số địa chỉ kết nối dữ liệu khác một nơi xuất phát rất tốt khác là trang dữ liệu Social Policy Virtual Library <http://users.utu.fi/thepap/SPVL/dpdbas.htm>.

Có hàng chục địa chỉ về chủ đề đặc biệt. Một trong số đó là Multiple Imputation FAQ page, ở <http://www.stat.psu.edu/~jls/mifaq.html>. Multiple imputation là một phương pháp điền các dữ liệu còn thiếu bằng cách sử dụng các biến khác để đoán các giá trị

còn thiếu. Một địa chỉ khác là một trang của Rubin giải thích propensity score analysis, ở

<http://www.symposion.com/nrccs/rubin.htm>. Propensity score analysis là một phương pháp giải quyết độ chệch tự chọn (self-selection bias). Ủy ban liên bang về phương pháp luận, ở

<http://www.fcsn.gov/spwptbco.html>, có một số báo cáo thú vị, đặc biệt là RL2. Record Linkage Techniques – 1997: Proceedings of an International Worrkshop and Exposition (Đây là RL2 không phải là RL1). Một địa chỉ chủ đề đặc biệt thú vị khác là Multilevel modelling project ở <http://multilevel.ioe.ac.uk/>. Một địa chỉ về kho dữ liệu là kdnuggets ở <http://www.kdnuggets.com/> (một thư tin và các kết nối chung đến địa chỉ các kết nối). Một địa chỉ dự báo là Federal

Forecasters Conference, ở <http://nces.ed.gov/surveys/ffc/>. Các báo cáo ở hội nghị có thể tải xuống từ địa chỉ này. Một địa chỉ hữu ích khác là Statistical Data Collection and Processing ở <http://www.unece.org/stats/archive/02.02.e.htm>. Có thể tải xuống các báo cáo và các tài liệu làm việc.

Cuối cùng, nhiều giảng viên có các bài ghi ở lớp đưa lên mạng. Hai trong số đó là các ghi chép ở lớp về dự báo của Bob Nau ở <http://www.duke.edu/~rnau/411out00.html>, và Phân tích chuỗi thời gian và các kỹ thuật dự báo của Hossein Arsham, ở <http://obelia.jde.aca.mmu.ac.uk/resdesgn/arsham/opre330Forecast.htm> ■

**Hoàng Tích Giang**

(Theo ISI Newsletter. Vol.26, No. 1/2002)